DOCUMENT RESUME

ED 040 929 24 SP 003 966

AUTHOR Payne, David A.
TITLE Estimating Costs for Development of Candidate Performance Evaluation Procedures.
INSTITUTION Georgia Univ., Athens. Coll. of Education.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO GEM-Bull-69-7
BUREAU NO BR-8-9024
PUB DATE 69
CONTRACT OEC-0-8-089024-311(010)
NOTE 14p.; Phase 1, Elementary Teacher Education Model

EDRS PRICE EDRS Price MF-$0.25 HC-$0.80
DESCRIPTORS *Estimated Costs, Measurement Instruments, *Student Evaluation, Teacher Education
IDENTIFIERS Comprehensive Elementary Teacher Education Models

ABSTRACT

This paper contains cost unit tables and instructions for their use in estimating the total cost of evaluating a given instructional objective or group of objectives. Included is a list of analytical procedures to be followed in the development of any device to evaluate student performance, (e.g., a unit exam in child development or an attitude scale relating to instructional methods). Tables for estimating development costs (a dollar cost for ten items) include differential cost factors for the behavioral area sample, the level of complexity, the format of the device, and the stimulus source. Tables for method of administration, method of scoring, method of item and test analysis are also included. A 25-item bibliography contains selected references related to the development of specific types of evaluation devices. (JS)

ED040929

GEORGIA EDUCATIONAL MODELS
The University of Georgia
College of Education
Athens, Georgia 30601

Estimating Costs for Development of Candidate Performance Evaluation Procedures

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

GEM Bulletin 69-7

David A. Payne, Ph.D.

1969

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Charles E. Johnson TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Note: The material contained herein was perpared as a working paper for a report of an investigation conducted by the Georgia Educational Models, a research and development project, pursuant to a contract with the United States Department of Health, Education and Welfare, Office of Education, Bureau of Research. It may not be reproduced without permission.

SP003966

Development of Evaluation Procedures

The development of any evaluation device must involve deliberate analytical procedures. If one takes such an approach, the sequence of steps involved is approximately as follows:

1. Specify the ultimate goals of the educaticnal process.
2. Derive from these goals the portion of the system under study.
3. Specify these goals in terms of expected student behavior. If possible and relevant, specify the acceptable level of successful learning.

For the GEM Project the first three steps have already been accomplished. It is assumed that further clarification of objectives will take place as the overall project and its sub-parts become operational and implemented. Proper steps should be taken to:

4. Determine the relative emphasis or importance of various objectives, their content and their behaviors.
5. Select or develop appropriate situations that will elicit the desired behavior in the appropriate

content or environment, assuming the student has learned it.

6. Assemble a sample of such situations so that together they best represent the emphasis on content and behavior previously determined.
7. Provide for the recording of responses in a form that will facilitate scoring but that does not change the nature of the behavior elicited so that it is no longer a true sample or an accurate index of the behavior desired.
8. Establish scoring criteria and guides to provide objective and unbiased judgments.
9. Try out instruments in preliminary form.
10. Undertake a complete item analysis.
11. Revise the sample of situations on the basis of try-out information.
12. Analyze reliability, validity, and score distributions in accordance with purposes of score use.

The foregoing steps would be followed no matter what type of instrument or procedure was being developed. They would apply in devising a unit exam in child development or an attitude scale relating to instructional methods.

The following cost unit tables represent the <u>total</u> cost in evaluating a given objective or group of objectives. They include consideration not only of development costs, but also of costs related to administration, scoring, revision, and a student-examiner time investment factor. The units within each table represent differential cost factors for a group of ten items or stimuli. Specific directions precede each table. In general, the procedure involves totaling the unit weights derived from each of the seven tables and multiplying by a cost factor. At this point in time, the cost factor is .10. Multiplying by this factor will give a dollar cost for 10 items. This will be subject to change as the costs of materials and services increase.

Assumptions

The cost unit estimates in the tables which follow were based on the assumptions that:

1. The instrument development involved combined efforts of (a) an evaluation consultant, (b) a subject matter expert, and (c) a graduate student who would oversee duplication, administration and data processing.
2. The development phase involved approximately 100 examinees who had been instructed in the material.

3. The final instrument, device, or procedure will result from a refinement of an item pool approximately twice as large as the expected final product. If a 20-item unit exam is desired, then the development phase might begin with 40 items. There are obvious exceptions, e.g. behavior samples gathered through the use of video tape.

4. If time and funds permit cross-validation of procedures, it is suggested that the cost be estimated by considering again Tables 5-7 after initial development costs have been determined. Cross-validation costs can be handled separately or added to validation costs. The cost factor of .10 again applies.

5. Only one form of each procedure will be developed. If more forms are desired, then obviously the total cost need only be multiplied by the number of forms.

6. Costs were development costs only. Costs to routinely administer and apply the resulting devices need to take account of information in Tables 5 (administration) and 6 (scoring). To estimate application costs obtain weights from these two tables and multiply by .03.

Tables for Estimating Development Costs

Behavioral Area Sampled

It is assumed that the development of items for the affective area will be more difficult than those in either the cognitive or psychomotor areas. Identification of the

Table 1.
Weights for Behavioral Area Sampled

Cognitive	Affective	Psychomotor
10	25	15

appropriate weighting factor relative to the type of item will of course be determined by the nature of the objective. In the majority of cases many items will be developed. Total cost of the device will be determined after determining the cost of a group of ten, i.e. after going through all seven tables.

Level of Complexity

Cost of item and instrument development should obviously be tied to the degree of refinement, complexity, and difficulty required. What is really being said here is that some items or procedures are more costly to develop than others. Basically, the lower category refers to knowledge (recall) and comprehension cognitive outcomes, attending and responding affective outcomes, and simple

psychomotor skills. The higher category includes application through evaluation cognitive outcomes, valuing, organization and characterization affective outcomes, and complex psychomotor skills. Simply add this weight to that selected from Table 1.

Table 2
Level of Behavioral Complexity

Lower	Higher
10	20

Format of Device

In attempting to evaluate a variety of objectives, one must of necessity employ a variety of techniques. Most of the frequently-used techniques are listed in Table 3. They range in degree of sophistication from simple straightforward rating scales to complex and refined scaled devices which employ during their development methods such as the method of paired comparison, equal appearing intervals, scalogram analysis and successive intervals. See Appendix A for a brief bibliograph concerning representative types of devices. Basically, these categories relate to the method of recording the examinee responses.

Table 3.
Format of Device

Behavioral Sample	15	Free Response	1	Rating Scale	3
Check-List	1.5	Observation Schedule	2.5	Scaled Device	10
				Semantic Differential	3
Forced-Choice	2.5	Opinionaire	2.5	Standardized Test	1

Stimulus Source

A cost differential factor should be taken into account as the development of a "new" item is considerably more expensive than pulling one from an old test or item file or modifying a previously-used item.

Table 4.
Stimulus Source

New	Adapted/Modified	Old
80	55	40

Method of Administration

A variety of methods are available for administering the procedure, both in trial form and in its polished state. The cost involved in using a trained examiner in a one-to-one situation can be substantial. Computer costs are those basically involved in initially establishing a retrieval system.

Table 5.
Method of Administration

Self-Group	Computer	Individual Administration
2.5	3.5	8

Scoring

Again the use of hand scoring methods, particularly those involving content analyses, analytical ratings and the like, can be quite costly. In most cases, several scorers should be used so as to help insure some objectivity.

Table 6.
Method of Scoring

Hand (Short Answer)	Hand (Extended Response)	Machine
4	10	2.5

Item and Test Analysis

During the development stage, routine examinations of test and item discrimination, validity, difficulty, and reliability should be undertaken.

Table 7.
Method of Item and Test Analysis

IBM 1230	Computer (New Program)	Computer (Library Program)
2.5	25	5

There are obviously several other factors that should be taken into account in estimating development and application costs. Prominent amoung these are time for student response, duplicating of instruments, and revision of instrument based on tryout data. It was felt that either it was impossible to estimate these costs or that they functioned as basically constant factors and were therefore treated as a lump-sum constant in the cost factor. A comment on one remaining factor should be made. It is almost always desirable to attempt an external validation of any measuring device. The necessity of gathering criterion data significantly increases development costs. A more effective instrument or technique will of course be the result. The determination of an external criteria might involve as much development effort (time, money, etc.) as did the construction, derivation, or modification of the original device. The overall cost obviously then doubles.

Illustrative Estimate

Let us assume that an instructor desires to estimate the cost of producing a 45-item single concept exam. The test will cover exclusively cognitive outcomes, measure only lower level outcomes, and have a multiple-choice format. In addition, the test will be group administered,

machine scored and analyzed. Using Tables 1-7, we determine the following weights:

Table		Category	Weight
1.	Behavioral Area Sampled	Cognitive	10
2.	Level of Behavioral Complexity	Lower	10
3.	Format of Device	Forced-Choice	2.5
4.	Stimulus Source	Old	40
5.	Method of Administration	Group	2.5
6.	Method of Scoring	Machine	2.5
7.	Method of Item and Test Analysis	IBM 1230	2.5
			70.0

Multiplying by a cost factor of .10 results in an estimate of $7.00 for 10 items of the type specified. The total cost for 45 items would be $31.50. One would probably need as many as 65-70 items in order to end up with the final 45 refined items. Cost adjustments obviously need to be made. As other factors come into play, costs will be influenced. For example, if instead of using old items, new ones were to be constructed, costs would increase. If the device were to include both lower and higher complexity items, costs would go up. Due to the fact that the objectives for GEM have been so well specified, requirements can be anticipated.

Selected References Relating to Development of Specific Types of Evaluation Devices

Check-Lists

Ahmann, J.S., & Glock, M.D. Evaluating pupil growth. (Third Edition). Boston: Allyn and Bacon, 1967.

Brandfield, J. M., & Maredock, H. S. Measurement and evaluation in education. New York: Macmillan, 1957.

Guion, R. M. Personnel testing. New York: McGraw Hill, 1965.

Forced-Choice

Osburn, H. G., et. al. The relative validity of forced-choice and single stimulus self-description items. Educational and Psychological Measurement, 1954, 14, 407-417.

Payne, D. A. The specification and measurement of learning outcomes. Waltham, Mass.: Blaisdell, 1968

Richardson, M. W. Forced-choice performance reports: a modern merit-rating method. Personnel, 1949, 26, 205-212.

Observation Schedules

Heyns, R. and Lippitt. Systematic observational techniques. In G. Lindzey (Ed.) Handbook of Social Psychology. Vol. 1. Cambridge: Addisen-Wesley, 1954, 370-404.

Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart and Winston, 1965.

Webb, E. J., et. al. Unobtrusive measures: nonreactive research in the social sciences. Chicago: Rand McNally and Company, 1966.

Opinionaire

Oppenheim, A. N. Questionnaire design and attitude measurement. New York: Basic Books, 1966.

Payne, S. L. The art of asking questions. Princeton, N. J.: Princeton University Press, 1954.

Rating Scales

Guilford, J. P. Psychometric methods. New York: McGraw Hill, 1954.

Guion, R. M. Personnel testing. New York: McGraw Hill, 1965.

Scaled Devices

Edwards, A. L. Techniques of attitude scale construction. New York: Appleton-Century-Crofts, Inc., 1957.

Semantic Differential Technique

Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart and Winston, 1965.

Osgood, C., et. al. The measurement of meaning. Urbana, Illinois: University of Illinois Press, 1957.

Snider, J. G. and Osgood, C. E. (Eds.) Semantic differential technique. Chicago: Aldine Publishing Company, 1969.

Standardized Tests

Buros, O. R. The sixth mental measurements yearbook. Highland Park, N. J.: The Gryphon Press, 1965.

Cronbach, L. J. Essentials of psychological testing, (Third Edition). New York: Harper and Row, 1970.

Mehrens, W., & Lehmann, I. J. Standardized tests in education. New York: Holt, Rinehart and Winston, 1969.

Miscellaneous References

Bonjean, C. M., et. al. Sociological measurement: an inventory of scales and studies. San Francisco: Chandler Publishing Company, 1967.

Bonney, M. E., & Hampleman, R. S. Personal social evaluation techniques. New York: Center for Applied Research in Education, Inc., 1962.

Ryans, D. G., & Frederiksen, N. Performance tests of educational achievement. In E. F. Lindquist (Ed.) Educational Measurement. Washington, D.C.: A.C.E., 1951

Shaw, M. E. and Wright, J. M. Scales for the measurement of attitudes. New York: McGraw Hill, 1967.

Swain, E. I. Evaluation and the work of the teacher. Belmont, California: Wadsworth, 1969.